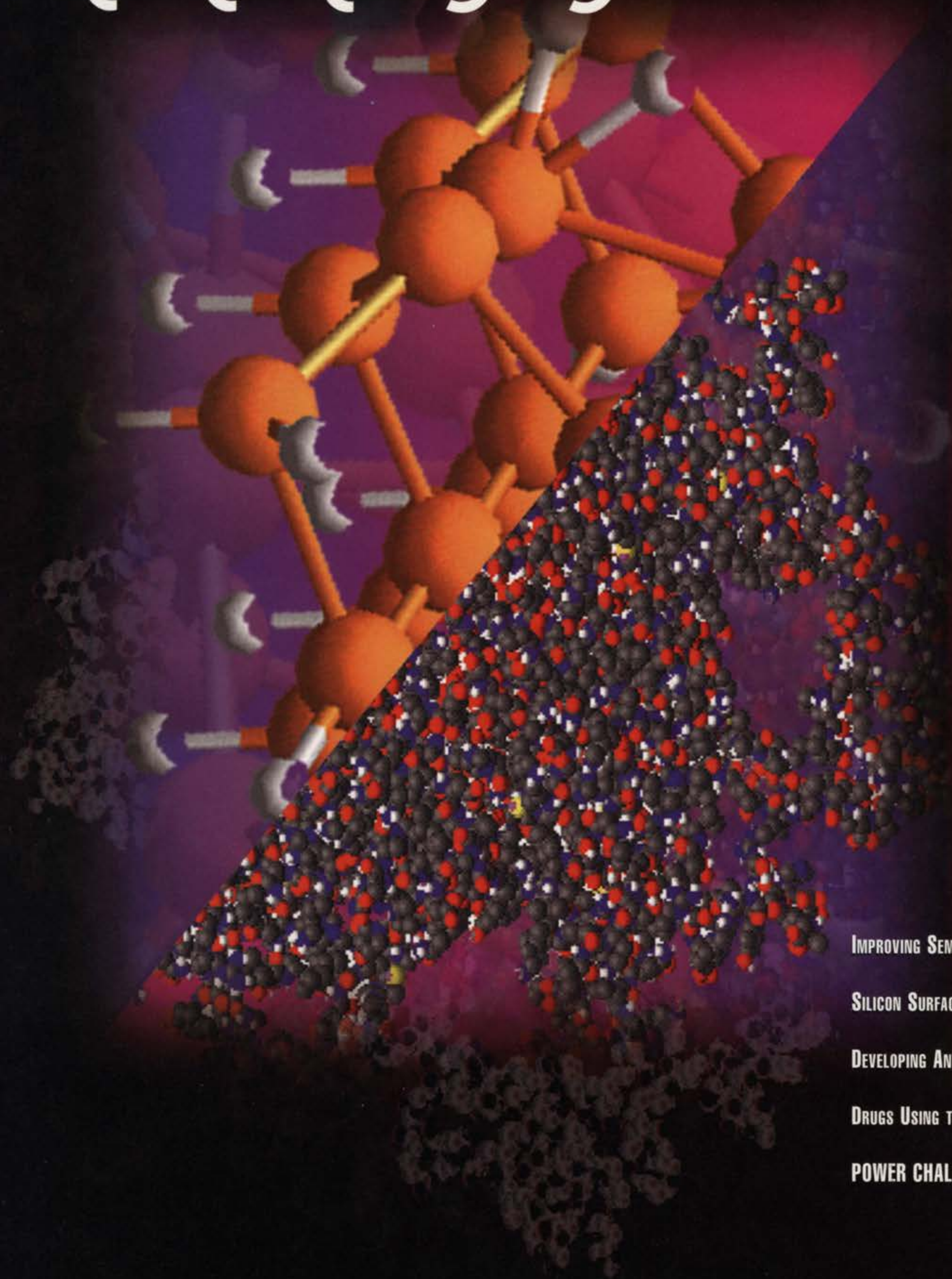
NCSA
VOL. 10 No. 1
SPRING 1996

access



IMPROVING SEMICONDUCTOR SILICON SURFACES AND DEVELOPING ANTIVIRAL DRUGS USING THE SGI POWER CHALLENGE SYSTEM

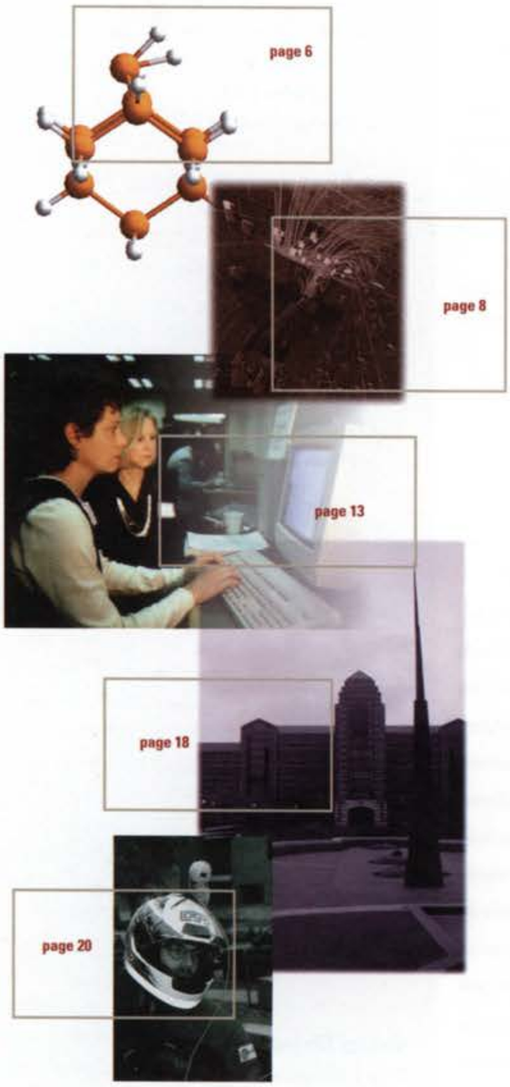
NCSA™

A TEN-YEAR FOUNDATION FOR THE FUTURE • UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

access

NCSA's high-performance computing magazine

DEPARTMENTS



RESEARCH

SPECIAL REPORT

OUTREACH + EDUCATION

NEW TECHNOLOGY

INDUSTRIAL PROGRAM

An apology

In the Fall 1995 *access*, one of the subjects of the photo on page 35 (group photo, lower right) was incorrectly identified. The woman pictured second from right was Tess Moon, University of Texas, not Nancy Sottos, UIUC. We regret the error.

NCSA Contacts Directory

NCSA Receptionist
(217) 244-0072
(217) 244-1987 FAX

Applications Group/Faculty Program
Melanie Loots (Associate Director)
(217) 244-2921
mloots@ncsa.uiuc.edu

Visitors Program
Jean Soliday
(217) 244-1972
jsoliday@ncsa.uiuc.edu

Computing & Communications
Charles Catlett (Associate Director)
(217) 333-1163
catlett@ncsa.uiuc.edu

Ken Sartain (Information)
(217) 244-0103
sartain@ncsa.uiuc.edu

Consulting Services
John Towns
(217) 244-1144
8:30 a.m.–5:00 p.m. CST
consult@ncsa.uiuc.edu

Networking
network@ncsa.uiuc.edu

Technology Management & Operations
Jeff Rosendale
(217) 244-0709
jeffr@ncsa.uiuc.edu
(217) 244-0710 (Services/Help)
help@ncsa.uiuc.edu

Education & Outreach Program
John Ziebarth (Associate Director)
(217) 244-1961
ziebarth@ncsa.uiuc.edu

Community Outreach
Scott Lathrop
(217) 244-1099
scott@ncsa.uiuc.edu

Information Technology Projects
Alan Craig
(217) 244-1988
acraig@ncsa.uiuc.edu

K–12 and Lifelong Training
Melissa Kelly
(217) 244-5606
mkelly@ncsa.uiuc.edu

Training Program
Beth Richardson
(217) 333-6276
ncsa-train@ncsa.uiuc.edu

Industrial Program
John Stevenson (Corporate Officer)
(217) 244-0474
johns@ncsa.uiuc.edu

Marketing Communications
Maxine Brown (Associate Director)
(217) 244-7255
maxine@ncsa.uiuc.edu

Allocations
Radha Nandkumar
(217) 244-0650
allocations@ncsa.uiuc.edu

Media Technology Resources
Tony Baylis
(217) 244-1990
tbaylis@ncsa.uiuc.edu
media@ncsa.uiuc.edu (Services)

Orders for Publications, NCSA Software, and Multimedia
Debbie Shirley
(217) 244-4130
orders@ncsa.uiuc.edu

Public Information Office
John Melchi
(217) 244-3049
(217) 244-8195 FAX
jmelchi@ncsa.uiuc.edu

Publications Group
Melissa Johnson
(217) 244-0645
melissaj@ncsa.uiuc.edu

Software Development Group
Joseph Hardin (Associate Director)
(217) 244-0409
hardin@ncsa.uiuc.edu

Jae Allen (Information)
(217) 244-7802
jallen@ncsa.uiuc.edu

Staff Resource Center
Recruiting, staffing, employment
(217) 244-0634
strc@ncsa.uiuc.edu

Virtual Environments Group
Tom DeFanti (Associate Director)
(217) 244-3408
defanti@ncsa.uiuc.edu

access

NCSA's high-performance computing magazine

Spring 1996

Managing Editor
Melissa LaBorg Johnson
melissaj@ncsa.uiuc.edu

Editor
Fran Bond
fbond@ncsa.uiuc.edu

New Technologies Editor
Ginny Hudak-David
virginia@ncsa.uiuc.edu

Copy Editors
Ginny Hudak-David
Paulette Sancken

Designer
Robb Springfield

Cover Design
Carlton Bruett

Printer
University of Illinois
Printing Division

NCSA

access (ISSN 1064-9409) is published by the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign (UIUC) with support from the National Science Foundation, the Advanced Research Projects Agency, other federal agencies, corporate partners, the University of Illinois, and the State of Illinois. Permission to reprint any item in *access* is freely given, provided that the author and *access* are acknowledged.

Copyright © 1996 Board of Trustees of the University of Illinois. All brand and product names are trademarks or registered trademarks of their respective holders.

PRINTED WITH SOY INK

PRINTED ON RECYCLED PAPER

This publication is printed on Evergreen Matte and Evergreen Matte Cover.

"Science is facts; just as houses are made of stones, so is science made of facts; but a pile of stones is not a house and a collection of facts is not necessarily science."

—Henri Poincaré (1854–1912)

Poincaré's thought-provoking words stimulate the imagination to focus on the objectives of scientific research. As he says, science is much more than "a collection of facts." Interpretation and analysis of facts bear fruit as well. NCSA's powerful resources in high-performance computing enable scientists to go beyond data gathering so they can concentrate on the creative aspects of their endeavor.

"Modeling Hierarchies—Doing More with Less" (page 4) and "Chemistry 'in silico' " (page 6) describe the work of scientists, ranging from atomic-level research in viruses to materials, who use NCSA's state-of-the-art HPCC systems. In "Riding the WAVE: I-WAY in Real Time at SC'95," you are updated on this prototype for distributed computing here and abroad (see pages 8–11). Many recent outreach activities extended to educators and the community that involve them in Internet connectivity are reported on beginning with page 12.

Read about the second generation of Hierarchical Data Format (HDF)—a useful NCSA-developed tool for managing scientific data—in "HDF: The Next Generation" (page 14). To learn about recent and upcoming Web-related activities turn to page 17.

NCSA's newest industrial partners are announced on page 19. Activities of some of the people at NCSA (page 20) complete this issue.

Mailed with this issue of *access* is a poster commemorating NCSA's tenth anniversary. Created by the award-winning designer Jack Davis, the images reflect highlights of the center. Davis also designed NCSA's tenth anniversary logo (see inside front cover). Stephanie Drake, NCSA editor, coordinated the poster and researched its images.

Many at NCSA were saddened to learn of the sudden death of Yuri Rubinski, pictured on page 17. The words of Joseph Hardin, associate director of NCSA's Software Development Group, express their feelings: "Yuri was a great supporter of NCSA and the vision that we all share of universal empowerment through advanced technology. We worked together on the World Wide Web Conference Committee, where Yuri was a founding member and a voice of reason and compassion. We will miss Yuri very much."

–Fran Bond, Editor

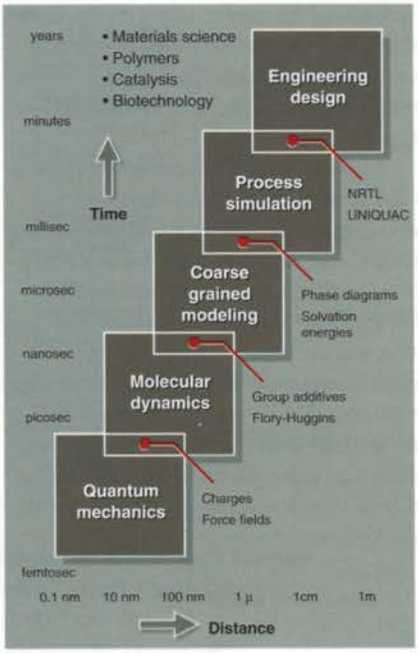
(Jules) Henri Poincaré, French mathematician, physicist, and author, was appointed to a chair of mathematical physics at the Sorbonne, Paris, in 1881 where he taught until his death. One of the greatest mathematicians of his time, he is remembered as the last great generalist in mathematics. Equally at ease in pure and applied math, Poincaré contributed significantly to analysis, algebra, topology, astronomy, and theoretical physics. He published considerably, including popular books on science for nonscientists. He originated algebraic topology and the theory of analytic functions of several complex variables. In his award-winning study of the three-body problem in celestial mechanics, Poincaré was the first to notice that small changes in initial conditions can produce chaotic behavior.

Modeling Hierarchies—Doing More with Less

Interview with Bill Goddard

Modeling all the atoms in the protein coat of the rhino virus is only one of Bill Goddard's goals. He has his sights set on bigger systems: larger viruses, antiviral drugs, and materials ranging from catalysts to superconductors.

by Holly Korab



Courtesy William Goddard III; illustration Marshall Greenberg

Hierarchical strategy for coupling atomistic simulations to engineering design.

Few scientists would disagree that theory should play a more central role in manufacturing. When William Goddard III says that within the next few decades quantum mechanics will drive industry, however, they may raise their eyebrows.

Quantum chemist Goddard, Ferkel Professor of Chemistry and Applied Physics at California Institute of Technology in Pasadena, CA, believes that faster computers and programs coupled with advances in atomistic simulations—based in quantum mechanics and molecular dynamics—are positioning quantum mechanics to move to the heart of manufacturing. It is a hefty claim.

Quantum mechanics deals with molecules at the subatomic level—calculating forces governing the behavior of electrons around a nucleus. These calculations are so massive that accurate simulations based on quantum mechanics are usually limited to 10 to 30 atoms. Given that engineering and technology deal with large systems—trillions upon trillions of atoms—applying quantum mechanics to industrial problems seems to many like driving a car by peering through a microscope. The scales are wrong.

Using hierarchical modeling

Goddard and other scientists who practice the "hierarchical" approach to modeling systems believe it is possible to link these scales. Hierarchical modeling progresses from the finest, most detailed level of understanding about a system—the level of quantum mechanics—to coarser representations by approximating interactions so that larger systems can be modeled for longer times. It broadens the scales by simulating bigger systems with fewer, faster calculations. Goddard's hierarchy has five levels (see chart below left). Although other researchers may have more steps, all culminate with engineering processes. "It is a telescoping whereby you can transcend from quantum mechanics to engineering with just a few levels of simulation," said Goddard.

Telescoping from one level of the hierarchy to the next is not easy. Each attempt at simplification increases the chances for inaccuracy in a simulation. Moving quantum mechanics simulations to those based on molecular dynamics demands that researchers establish boundary conditions and approximate the multitude of interactions in a system in ways that will not cause distortions. Often a methodology that works for one process may not work for another.

Because it is so difficult to simplify forces in order to broaden scales, Goddard has been moving back and forth between the lower levels of his hierarchy—quantum mechanics and molecular dynamics—for the past 30 years. He has focused on all five steps of his hierarchy in the last 6 years. Recently he and his team at Caltech's Materials and Process Simulation Center enjoyed a breakthrough that may propel them a step higher. His team developed several hierarchical modeling methodologies that increased the size, length, and time scales for molecular systems at least 1,000 times. As part of a study to understand how drugs help prevent the common cold, they simulated all 480,000 atoms in rhino virus 14, the cold virus. Now they are poised to expand to larger systems.

Last August Goddard and his colleagues ported their codes for simulating the virus, plus several others, to NCSA to optimize them on the SGI POWER CHALLENGEarray. The system's innate scalability attracted the researchers as much as did its long-term viability. Developing modeling methodologies is only half of the group's work. The other half is optimizing methodologies for high-performance computers they think researchers will use in the future. A look at how they are using their codes in the virus simulation provides insight into how they attack problems via the hierarchical approach to modeling.

What they are optimizing

The codes used in simulating the rhino virus 14 are two of many codes Goddard is now optimizing with NCSA's resources. Most of Goddard's work focuses on applications in materials science, often funded by industry; however, the virus work is Grand Challenge research funded by NSF. Although applicable to molecular biology, the program and its two underlying codes are being adapted for other uses, such as in industrial polymers and electronic materials.

The MPsim program, or Massively Parallel Simulator, was developed specifically by Goddard's group to simulate systems with greater than a million atoms. MPsim performs molecular dynamics calculations using two breakthroughs developed at Caltech: the Cell Multipole Method (CMM) for fast calculations of nonbonded interactions and the Newton-Euler Inverse Mass Operator Method (NEIMO) for calculating the response of a system to nonbonded forces.

CMM overcomes a bottleneck in large systems that arises from having to sum long-range Coulomb interactions. Coulomb forces are the electric charges between particles. Unlike other nonbonded interactions (van der Waals forces and hydrogen bonds), they do not fall off rapidly with distance. The electrostatic interaction between every particle must be calculated for all pairs of atoms. For a million-atom system, the number of calculations for Coulomb forces reaches into the trillions. Goddard's group adapted fast multipole strategies (originally done by Rokhlin and Greengard at Yale University) to large periodic systems using a hierarchy of meshes to reduce the number of interactions to be calculated from trillions to millions.

Testing the methodology

The rhino virus simulation presented an ideal test for CMM. Scientists know that the virus transmits its RNA to healthy cells by binding to receptors on the surface of the target cell. Once bound it releases its RNA through microscopic holes at the receptor sites. Scientists have speculated that drugs lessen or prevent colds by stiffening the walls of the receptor sites so that they cannot expand to accommodate the much larger viral RNA as it attempts to slither into the cell. Because stiffness is determined by the interactions of all the atoms in a protein coat, it was essential to simulate every atom in order to confirm the action of drugs.

To look at collections of viruses or other proteins, Goddard's group will need to enlarge their simulation by coarse graining the virus structure. Coarse graining refers to looking at collections of atoms rather than individual atoms. Rather than simulating all

Ribbon display of the asymmetric structural unit (center) made up of four individual viral proteins (corners).

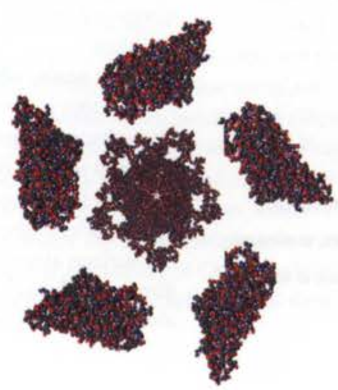
516,000 atoms of the protein coat, the group will employ the NEIMO tool to represent only the virus's 240 globular proteins as independent entities.

NEIMO was first developed by Abhinandan Jain at NASA and Caltech's Jet Propulsion Laboratory for guiding robotic arms in space satellites. It is a means of simplifying motion by restricting the degrees of freedom within a system. (Jain developed a way to calculate the relationship of the matrices between movement and torque–the fundamental physics behind these ordinary gestures–that scales linearly rather than to the third power.) Nagarajan Vaidehi of Goddard's group adapted this methodology to proteins so that the bulk of a protein can be treated as rigid and only some of the surface elements as flexible. In their simulation of the rhino virus, the team will ignore the small changes at the bonds and angles within its 240 proteins and calculate only the degrees of freedom at those bonds that work like door hinges.

Filling in the gaps

Simultaneous with these developments, Goddard's group has been backtracking to develop more accurate force fields, which will be required for drug design. The force fields available to researchers in this area are nearly a decade old. While adequate for studying drug effects, they may not provide the detailed information critical for predicting molecular interactions. Since Goddard's group hopes to move further into drug design, they are collaborating with Richard Friesner of Columbia University to develop a new quantum mechanics code. The new code would enable them to include solvents while handling very large molecules.

This process is not a concern to Goddard who knows that climbing the modeling hierarchy often requires taking a step back. What is important is keeping one's eyes on the target, he said. "The whole point," said Goddard, " is that theory can solve real problems. In the past there have been gaps between fundamental theory and manufacturing, but, gradually, we are filling in those gaps."



Space-filled model of a pentagonal section of the rhino virus surface made up of five asymmetric structural units.

Images courtesy William Goddard III

Holly Korab is a science writer in the NCSA Publications Group.

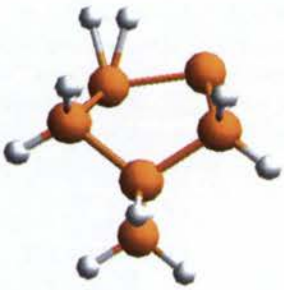
Materials and Process Simulation Center, Caltech http://www.wag.caltech.edu/

Chemistry "in silico"

Interview with Ken Jordan

Silicon is the most widely used electronic semiconductor material, and it has been for over three decades. It is an integral component of everything from transistor radios to communications satellites, pocket calculators to supercomputers.

by Sara Latta



Model used by Ken Jordan and colleagues to study desorption, or removal, of hydrogen from the surface of Si(100).

Yet despite its ubiquity, certain aspects of the surface chemistry of silicon (Si) crystals remain a mystery. A particularly hot area of research is the mechanism of desorption (removal) of hydrogen from the surface of Si(100), a critical step in semiconductor manufacture. The measured kinetics of hydrogen desorption are not as expected, which sends many theorists back to the drawing board—or the computer.

One of those theorists is Ken Jordan, professor of chemistry at the University of Pittsburgh. Jordan and his collaborators are using NCSA's SGI POWER CHALLENGEarray system, along with the Pittsburgh Supercomputing Center's (PSC's) CRAY C90 computer, to construct Si(100) surface chemistry models that may resolve the problem.

On again, off again

Si(100) is the most important surface in the semiconductor industry. The numbers 1, 0, and 0 refer to the crystallographic orientation of the silicon atoms with respect to the surface. In the most stable arrangement, the surface atoms are arranged in pairs, or dimers, with each atom of each dimer pair also bonded to two silicon atoms in the second layer. The process of slicing the silicon crystal into thin, round wafers leaves unpaired electrons, resulting in a chemically reactive surface. "You couldn't have such a reactive surface being used in devices," said Jordan, "because oxygen and water vapor in the atmosphere would react with the surface and change its properties in an unpredictable manner."

Scientists usually deal with this problem by flooding the newly cut silicon surface with hydrogen gas to tie up the unpaired electrons. This gives rise to a stable surface. In order to make useful semiconductor devices, one must first etch the surface or deposit more silicon atoms. In either case it is necessary to displace those hydrogen atoms, usually by heating the surface. The displacement of hydrogen from the surface sounds fairly straightforward, but the mechanism for this process has turned out to be surprisingly controversial. One might think that each hydrogen-silicon bond would be broken directly, leaving hydrogen atoms free to form H_2 (hydrogen gas). But that requires a lot more energy than is actually used; it is clearly not the correct mechanism.

A more likely process involves an intermediate species with two adjacent hydrogen atoms partially bonded to each other and to one silicon atom. To test this theory, Jordan created a computer cluster model—a minisilicon surface consisting of nine silicon atoms, the two surface hydrogen atoms, and 12 additional hydrogen atoms to properly terminate the subsurface silicon atoms.

Jordan used a variant of the widely used density functional theory, or DFT, to

Images courtesy Ken Jordan

describe the silicon-hydrogen interactions. (DFT calculates the total energy of the system from the electron density and is less computationally demanding than methods that are based on wavefunctions.) While density functional methods are not in general known for their accuracy in describing hydrogen bonding, Jordan used PSC's CRAY C90 system to compare a variety of density functional methods to a highly accurate many-body method. He found one method, or functional as they are called, to be particularly well suited to describing the interactions between the silicon surface and hydrogen bonds: Becke3-LYP.

Applying the Becke3-LYP functional to the Si_9H_{14} clusters, Jordan "found that the activation energy for going through the intermediate is appreciably higher than experiment, although other researchers using other models and theoretical methods get a lower energy for this process."

To counter criticism that the Si_9H_{14} clusters were too small to reflect the behavior of a silicon surface, Jordan and coworkers have recently adopted more realistic models containing up to 21 silicon atoms. "So far," Jordan said, "the energy barrier for H_2 desorption does decrease with adoption of larger cluster models, but not by enough to be consistent with experiment."

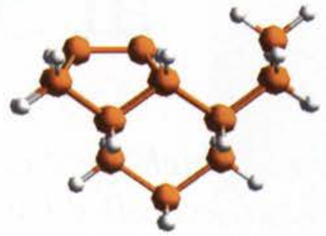
Not surprisingly the 21-atom model necessitated an increase in computing power. "We had really gone about as far as we could on the small clusters on our workstations," said Jordan. "Having access to the POWER CHALLENGE systems at NCSA has made feasible extending the calculations to these larger clusters." Jordan estimates that he has used 2,000 service units on NCSA's POWER CHALLENGE over the past year. Most of the cluster model calculations were carried out using the Gaussian 94 program. "The Gaussian 94 DFT code is well suited for the POWER CHALLENGE," said Jordan. "We get good speedup on up to six processors."

Recently Jordan's group and two others have proposed that the mechanism for hydrogen removal from the surface may involve defects. Surface defects come in a variety of types. Jordan and coworkers have focused their attention on defects involving unpaired surface silicon atoms. "One might expect the hydrogen atoms to move around the surface until 'captured' by the defects," said Jordan. "We have established the existence of a lower energy path that works through a bond-shifting mechanism; the net result is that it looks like the defect has moved."

Other approaches

Jordan is not alone in his search for better methods to uncover the mechanism for hydrogen removal. Two other research groups, using NCSA's resources, are searching for the elusive mechanism.

Richard Martin, UIUC physics professor and coworkers, have recently used NCSA's POWER CHALLENGE and TMC's CM-5 systems to model giant fullerenes containing as many as 3,000 carbon atoms and a DNA molecule segment, using density functional methods. Martin is also developing "Order N" or linear scaling methods in which the computational time increases linearly with the size of the system. Martin notes that present methods scale as a power, usually a cube, of the size of the system. He is also evaluating a number of density functional methods for their efficacy in modeling silicon surface defects and the interactions between hydrogen and silicon surfaces. The latter project, Martin stressed, is still in the development stage. Workstations were used to study small cluster models before scaling up to NCSA's POWER CHALLENGE, which is Martin's machine of choice.



Models for two states of desorption (reading from top to bottom): Final transition state for H_2 desorption from a defect site; dihydride intermediate along the defect mechanism reaction path.

Jeffrey Grossman, UIUC graduate research assistant at NCSA, and Lubos Mitas, UIUC postdoctoral research associate at NCSA, while not studying the interaction between hydrogen and silicon clusters *per se*, are using a Quantum Monte Carlo (QMC) method to study large silicon and carbon structures. QMC methods, initiated by (among others) David Ceperley, UIUC physics professor and NCSA research scientist/leader of the Quantum Physics Group, employ a stochastic approach to solve the many-body Schrödinger equation with a high degree of accuracy. It is a more general approach that, while computationally expensive, can be applied to a wide range of systems. Grossman and Mitas have shown that a QMC method can be applied equally well to systems ranging from one atom to clusters of 20 atoms, even to solids—essentially an infinite system. "Usually," Mitas said, "very different methods are required to study such a range of systems." Grossman and Mitas used PSC's CRAY C90 system and NCSA's SGI POWER CHALLENGE and HP/CONVEX Exemplar systems. Martin notes that QMC methods might be useful to calibrate the accuracy of less computationally demanding but more specialized density functional approaches.

Matters of practicality

Studying the removal of hydrogen from silicon surfaces has great practical importance. "Suppose that H_2 desorption really proceeds through defects as we proposed," said Jordan. "Then if one could remove or 'block' those defects, H_2 removal would have to go through a higher energy process, and the surface would be more stable at elevated temperatures. This could be important in some applications."

There is another, more fundamental reason for pursuing the elusive H_2 desorption mechanism. "It serves as a testing ground for various theoretical methods," said Jordan. "Before extending our studies to more complicated surface processes, we want to be sure that the methods we are using are 'up to the task'."

Jordan's collaborators on silicon surfaces are Petr Nachtigall, Czech Academy of Sciences in Prague, Czech Republic; Carlos Sosa, Cray Research Inc.; and Hannes Jonsson and Arthur Smith, University of Washington. Martin's colleagues on developing DFT methods are Pablo Ordejon, former postdoctoral research associate now at University of Orviedo, Spain, and Satoshi Itoh, former UIUC visitor now at Hitachi Central Research Lab, Tokyo.

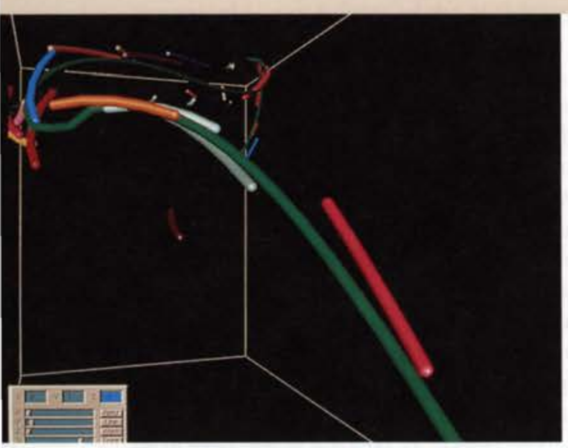
Sara Latta is a freelance science writer.

Riding the WAVE: I-WAY in Real Time at SC'95

by Holly Korab

In December 1995, the high-speed experimental distributed computing project called I-WAY (Information Wide Area Year) made its long-anticipated debut at SC'95. It was not flawless, but it propelled collaboration technology years ahead by presenting a workable prototype for distributed computing.

Ian Foster, ANL computer scientist, describes the software effort at SC'95. Foster led 30 software designers in developing I-WAY's distributed computing environment that includes Point of Presence (I-POP) machines and I-Soft software.



The Terabyte Challenge: High-Performance Computing and High-Performance Data Management Using the NSCP Meta-Cluster, **Robert Grossman, UIC, Dept. of Mathematics, Statistics, and Computer Science.**

Further information can be obtained in the catalog *Virtual Environments and Distributed Computing at SC'95: GII Testbed and HPC Challenge Applications on the I-WAY*, edited by Holly Korab and Maxine Brown, which is available from the Web at http://www.ncsa.uiuc.edu/General/Training/SC95/GII.HPCC.html.

Though not with the power of a tsunami, the next wave of high-performance networking that debuted at SC'95 in December proved, to anyone who still held doubts, that distributed metacomputing is on its way. During three-and-a-half frantic days of the conference, researchers and networkers demonstrated all the essential functions for distributed computing that they had built from scratch during the previous nine months. For instance they used a single interface to schedule and initiate runs on remote computing resources, and they used Message Passing Interface (MPI)—a software library for porting applications to other computer architectures—to run applications in heterogeneous computing environments. They accomplished these by winning the support and cooperation of telecommunications carriers and 30 different academic and research institutions [see *access*, Fall 1995, page 14].

In his keynote address, William A. Wulf, professor of engineering and applied science at the University of Virginia and former assistant director of CISE, crowned I-WAY "a national treasure." Another person likened it to inventing the internal combustion engine, automobile, and interstate highway system all at once.

"We wanted to show that we could push the concept of high-speed ATM networks across the country and create the framework that would make this happen–all at the same time," I-WAY cofounder Tom DeFanti said. "We knew it was possible, but others said it was not. We built the pieces, and this was the test. And it worked." DeFanti is associate director of the Virtual Environments Group at NCSA and director of the Electronic Visualization Laboratory at UIC.

Lessons learned

Exactly what worked—and particularly what did not work—will be dissected and researched this next year. More than just a demonstration, I-WAY was a testbed with two goals: to present a prototype for high-speed distributed computing and to identify its weaknesses.

Distributed computing links remote visualization and computing resources via high-speed networks into single virtual computing systems. Because this schema is a faster, more realistic means of assembling computing power than is trying to amass it in one place, many people believe distributed computing is the future of high-performance computing. The problem is that high-speed networks are not interoperable. Like the railroads of the 19th century that were built to different gauges, high-speed networks are of varying protocols and routing and switching technologies. I-WAY was the "first draft" of standard gauges and switches for distributed computing.

"I've always viewed I-WAY as a process, not a thing," said Gary Minden, program manager for the Advanced Research Projects Agency's (ARPA) Information Technology Office and a principal supporter of I-WAY. "Its purpose is to find out what the right questions are."

Although the right questions will be winnowed out during a follow-up meeting being arranged by Minden and representatives from NSF—I-WAY's other primary backer—a few of the likely topics already can be gleaned from the experiences at SC'95.

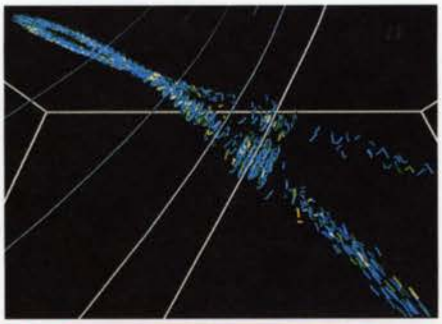
A topic likely to appear at the top of the list, because it embraced so many aspects of the project, is promoting better collaboration among all layers and entities of a distributed system. Of the 60 2D and 3D large-scale applications scheduled to run over the I-WAY (jointly called the GII Testbed and HPC Challenge events), only about half did so. Some researchers changed their plans and ran the applications locally using precomputed data. Of those who attempted to use the I-WAY, a handful were forced to resort to precomputed data or run their applications on the Internet. Many of their problems were traced to glitches in their applications. Others were due to inconsistencies elsewhere in the network, such as the incorrectly configured router at one institution that caused one researcher's application to creep across the I-WAY at only 5% of its anticipated bandwidth. More often than not, problems with the I-WAY stemmed from fledgling interconnections between the networks—pieces of which had been assembled in less than nine months and cobbled together in a week.

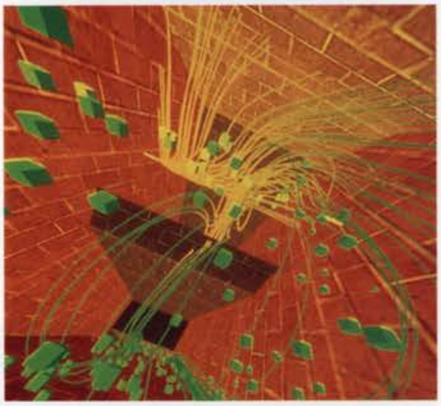
The amount of information required for engineering the final stage of network assembly was enormous. Many of the configurations were complicated by the need to access remote resources from showfloor booths as well as from the GII Testbed. Many of these requirements were uncovered once the WAVE team assembled in San Diego. (WAVE—short for Wide Area Visualization Experimental—was the name given to the portion of the I-WAY inside the convention hall.) With time at a premium and complications mounting, the vBNS (very high-speed Backbone Network Services) was the only one of 10 networks operating at the start of the GII Testbed. The other networks came up over the next three days; eventually all carried GII applications. Still the vBNS was used most extensively.

That the vBNS worked as well as it did owed much to collaborations, said Linda Winkler, a computer scientist at Argonne National Laboratory (ANL) and one of the primary architects of the I-WAY network. She is convinced vBNS would not have been running by the first day of the GII Testbed if specialists from the NSF supercomputing centers and MCI had not pitched in. "MCI made sure we understood the connectivity, and the centers made sure their resources were accessible. When two of the routers weren't responding, they all poked their heads into the problem to solve it," said Winkler.

Collaboration among 30 software designers led by Ian Foster at ANL produced one of the most creative components of I-WAY. The I-WAY Point-of-Presence (I-POP) machines and I-Soft software that comprised the I-WAY distributed computing environment shielded researchers from many of the intricacies of the networks. Researchers in the GII Testbed would log into any one of 17 I-POPs, schedule time on "I-WAY virtual machines," and initiate runs on required resources without being concerned with such issues as authorization, scheduling, and network interfaces at each remote site. That was what happened when everything worked as planned. As with the networks, circumstances often conspired to complicate this scenario. With application requirements and network connections changing hourly, Foster's team often found themselves manually reconfiguring virtual machines and network interfaces as well as chasing bugs in user programs and I-Soft software.

These kinds of on-the-fly changes had their benefits, such as an improvement to the I-Soft MPI developed for the I-WAY Asynchronous Transfer Mode (ATM) networks. NCSA astrophysicist Ed Seidel and NCSA computational biologist Marcus Wagner both used the specialized MPI. In so doing, Seidel discovered two MPI bugs, which Foster's team corrected in time for Wagner's demonstration.





***Interactive Scientific Exploration of Gyrofluid Tokamak Turbulence*, Gary Kerbel, Lawrence Livermore National Laboratory, National Energy Research Supercomputer Center (above); *Remote Engineering Using CAVE-to-CAVE Communications*, Lori Freitag, Argonne National Laboratory**

Interplays like those between Seidel, Foster, and Wagner—which produced a better MPI—attest to why collaboration is essential for development of distributed computing. Networkers need feedback. "Researchers will have to delve more deeply into the workings of these networks if their applications are to perform successfully," said Rick Stevens, director of the Math and Computer Science Division and Computing and Communications Center at ANL, and who, along with NCSA Director Larry Smarr, is the third cofounder of I-WAY. "Researchers have to understand that the network is introducing a whole new set of variables into their application. It's like the adjustments they had to make when switching from vector to massively parallel programming—there are more things that can go wrong."

Another topic certain to dominate follow-up discussions about I-WAY is management tools to simplify interactions with the network. Foster envisions automated programming tools that will perform many of the I-POP tasks now done manually, such as scheduling resources and configuring virtual machines. "Eventually you want the applications people to specify the computing resources they need and have the auto-

(Continued on page 10)

mated scheduler reserve time and resources rather than requiring the researchers to select the machines and capabilities they want, which was an interim measure we adopted for I-WAY," said Foster.

Rémy Evard would like to see tools developed for ATM, the transfer protocol being tested on the I-WAY. He spent hours manually debugging the ATM-based network. Evard, director of technology at Northeastern University who was team leader for WAVE, likened debugging ATM to editing a book written on index cards. "Imagine if someone can only give you a yes or no answer as to whether they like your book. If they answer no, then you have to go back through the stack of cards one-by-one to discover where the error lies. If a conjunction were missing and you put it in, suddenly the book works. That's debugging with ATM. If in the middle of all this a piece of hardware goes down, that is like someone flicking 10 cards out of the middle of the stack," said Evard.

I-WAY heroes

In addition to teaching everyone a lot about constructing a distributed system, I-WAY generated its own brand of legends. Few people operating behind the scenes will forget the researchers and programmers who debugged all day and all night. These people earned I-WAY hero reputations for their endurance. Others acquired reputations for grace under

Behind the scenes at SC'95, technical staff set up the CAVE for the GII Testbed exhibits.

Graduate Research Assistant Daniel Weber (left) and NCSA Research Scientist Ed Seidel (right) debug their gravitational wave application.

pressure. When the network finally came up at 10:05 a.m. on the opening day of the conference, researchers scheduled for the first three slots in the GII Testbed had only one hour and 50 minutes to test and debug their applications. They did it.

Special praise is due those responsible for networking during the conference. In addition to I-WAY, they integrated three networks: SCinet (the onsite production network), a wireless local area network, and WAVE. Strung through the rafters were 150 miles of fiber optic cables with 840 tails (or connections). Connectivity within the convention center was compared to that of a small city.

Networking crews began arriving the week before the conference. These were the tactical assault units from the five supercomputing centers as well as ANL and Northeastern University. As the week progressed, the population inside the Plexiglas-enclosed network operations center swelled to about 50 or 75 and included additional assistance, such as two network engineers from the National Institutes of Health and six students from the Naval Postgraduate School in Monterey, who sacrificed studying for their final exams to participate in marathon networking.

Twenty-hour days were the norm. Flexibility was the key as they bandaged connections and tried to satisfy ever-changing requirements and expectations. Evard summed up the experience as "10 days of nonstop problem solving and knowing exactly what your priorities were."

Birds of a feather

Setting priorities will be essential in continuing the work begun with I-WAY. During a birds-of-a-feather session at SC'95, participants emerged from the GII Testbed to indicate what they thought should be the next steps in distributed computing. On their list was further automation of the I-POPs as was the development of network management tools. Also mentioned was network-based collaborative software for designing and managing distributed computing networks.

A priority for NSF and ARPA was maintaining the enthusiasm for high-bandwidth distributed computing that was evoked by I-WAY. Enthusiasm will be essential if science is to transform the spurts of flawless networking witnessed at SC'95 into a solid entity. DeFanti believes that is a goal within science's grasp. He reminded the participants of his earlier experiences with virtual environments. When three CAVEs running 40 applications from researchers across the country were demonstrated at SIGGRAPH '94, those systems, like I-WAY, were beset with problems. This year they performed flawlessly. "The virtual environments came a long way in a year," said DeFanti. "In five years distributed computing will be commonplace, and you will look back on this event as a watershed in networking."

Holly Korab is a science writer in the NCSA Publications Group.

GII Testbed applications http://www.ncsa.uiuc.edu/General/Training/SC95/GII.HPCC.html

NCSA Advises

I-WAY applications pushed the limits of distributed computing at SC'95, and soon they will help set new boundaries for broadband networks.

NSF invited researchers with I-WAY applications to participate in an international broadband testbed as part of a new G7 effort to promote global interoperability for high-speed networks. The goal is to provide a medium for joint research and development. Initially that will mean establishing experimental intercontinental links between the U.S., Europe, and Japan.

Telecommunications companies have formed two consortiums that will donate test facilities for the project. Sprint—in collaboration with TeleGlobe (Canada), Deutsche Telekom (Germany), and France Telecom—will provide at least one 155

NSF on Global Broadband Testbed

Mbps line for a transatlantic testbed configured in a rectangle between the four countries. AT&T along with NTT (Japan) will provide a transpacific line of similar speed. Infrastructure for both testbeds is expected to be in place later this year. Eventually the United Kingdom and Italy—the remaining European G7 countries—will become part of the network.

U.S. participation in this project is being coordinated by Steve Goldstein, NSF program director for Interagency and International Networking Coordination. Advising Goldstein on how best to connect and use these networks are four members of NCSA. Charlie Catlett, associate director for Computing and Communications, and Randy Butler, manager of Networking and Security, are helping identify the best technologies for interconnecting U.S. networks with those in Europe and Japan. This project will involve interoperation of at least seven different broadband networks and exploration of interconnection strategies as well as mechanisms for dealing with delays inherent in multicontinent networks.

Helping channel some of the enthusiasm of the I-WAY to the global testbed are Tom DeFanti, associate director of NCSA's Virtual Environments Group and director of the Electronic Visualization Laboratory (EVL) at UIC, and Maxine Brown, associate director for Marketing Communications at NCSA and associate director of EVL. They are helping identify applications to run on these testbeds. Researchers who presented applications at SC'95 are being encouraged to expand the reach of their projects by collaborating with overseas colleagues. They are not the only researchers encouraged to participate in the project; anyone interested should send email to g7@ncsa.uiuc.edu.

The impetus behind what is now being called the Global Interoperability for Broadband Networks (GIBN) project was the Clinton Administration's push for a global information society. During the Naples Economic Summit in July 1994, President Clinton urged the G7 nations to develop an international information infrastructure. In response the G7 hosted the first Ministerial Conference on the Information Society in Brussels, Belgium, the following February. Eleven pilot project themes were identified there; among them was global interoperability of broadband networks. NCSA was one of 15 organizations invited by the U.S. Department of Commerce to exhibit information technologies at the conference [see *access*, Spring 1995, page 30]. NCSA's participation led to its involvement in the global testbeds.

The U.S. coordinator for the G7 Global Information Society initiative, which includes GIBN as one of its themes, is Thomas Kalil, National Economic Council, The White House.

G7 Global Information Society Initiative http://info.ic.gc.ca/G7

GIBN Applications http://www.ncsa.uiuc.edu/General/GIBN/

NCSAers Win Award at SC'95

by Allison Miller

***Galaxies Collide on the I-WAY: An Example of Wide-Area Collaborative Computing*, Mike Norman, UIUC/NCSA**

NCSA staff members Michael Norman, Greg Bryan, and John Shalf brought home the Third Annual High Performance Computing Challenge Award for Best Integration of Heterogeneous Applications from SC'95.

NCSA was one of four supercomputing centers and several research agencies sponsoring researchers in the Supercomputing '95 High Performance Computing Challenge. The contest featured 10 groups of researchers seeking to outdo each other in the race for the first teraflops application.

The goal of researchers participating in the HPC Challenge was to assemble the greatest amount of computing power and speed in running a single scientific application. A demonstration of the Grand Challenge Cosmology Consortium (GC^3), the winning entry was funded by NSF.

NCSA research scientist and GC^3 member Mike Norman, NCSA graduate research assistant Greg Bryan, and NCSA research programmer John Shalf used multi-supercomputing power in their simulation of galaxy collisions, titled "Galaxies Collide on the I-WAY: An Example of Wide-Area Collaborative Computing." Collaborators from other GC^3 institutions were John Dubinski and Lars Hernquist, University of California–Santa Cruz; Dennis Gannon, Kate Keahey, and Shelby Yang, Indiana University; and Joel Welling, Pittsburgh Supercomputing Center.

Allison Miller, a senior in rhetoric at UIUC, is a student intern in the NCSA Marketing Communications Group.

Galaxies Collide on the I-WAY http://zeus.ncsa.uiuc.edu:8080/GC3/SC95/challenge.html

Resource for Science Education Program Update

Compiled by Umesh Thakkar

The Resource for Science Education (RSE) Program, funded by NSF, has been gaining support and encouragement from the university and K–12 community (see *access*, Summer 1995, page 31). Members of the program include Scott Lathrop, Lisa Bievenue, and David Curtis.

The main objectives for the program are to (a) invite educators (students, teachers, and faculty) to form collaborations with NCSA/UIUC scientists and staff; (b) build bridges between research and classroom learning and teaching; and (c) empower teachers to improve their practice by integrating new information technologies into classroom activities.

Fall outreach

Fall 1995 activities in science education and new technologies outreach were as follows:

• RSE hosted Suzanne Damarin, professor of education at Ohio State University, whose interests are in virtual reality and mathematics education. Yasumasa Kanada, professor of information science at University of Tokyo, lectured on "Pi Calculation—Pushing the Limits of Number Crunching Machines" to the university and K–12 education community. Kanada's visit was co-hosted by Georgette Moore from the Yankee Ridge Elementary School, Urbana, IL.

• Program teachers from Yankee Ridge Elementary School, Urbana Middle School, and Urbana High School recently received funding from the Illinois State Board of Education for further development of online projects in pi mathematics, groundwater quality, and stock market activity that they started last summer. As one example the funding will support integration of geographical information systems (GIS) tools and data into Urbana Middle School's groundwater project.

• RSE was showcased at the Supercomputing '95 K–12 education poster session, and RSE teachers from Illinois, Nebraska, and Virginia shared their summer experiences (see *access*, Fall 1995, page 23) with other educators in panel and poster sessions.

• RSE staff assisted David Curtis, project director of NCSA's recently released Web site "Science for the Millenium" (see *access*, Fall 1995, page 25), in a preliminary evaluation of online science exhibits in astronomy and astrophysics by two groups of local high school teachers and students. A Science Expo was presented in the Education Program at Supercomputing '95 as part of a panel on "New Trends in Supercomputing for Classroom Use and Assessment." More than 70 educators signed up after the panel to further their interests in exploring how the Expo could be utilized in their classroom science activities. The panel presentation is accessible from the RSE Home Page. (See URL left.)

New program participants

In addition to several NCSA scientists and staff who have been interacting with RSE teachers since last summer, three UIUC faculty and staff members recently became involved with the program. They are as follows:

• Steven Hall, project director for UIUC Collaborative Visualization (CoVis), has begun collaborating with local RSE-affiliated elementary teachers who are interested in weather. These teachers use CoVis resources, such as the UIUC CoVis Geosciences Web Server and the Weather Visualizer, in their online weather curriculum projects. (See URL left.)

• Alfred Hubler, professor of physics and research scientist at the Center for Complex Systems Research, is utilizing CyberProf in Urbana High School's advanced physics classroom. An intelligent student-teacher interface that facilitates grading, creating, and presenting educational course materials, CyberProf was developed by Hubler's research group. (See URL left.)

• Elisabeth Oltheten, professor of finance in the College of Commerce and Business Administration, is integrating the University of Illinois Securities Exchange Simulation (UISES) into stock market projects for middle and high school classrooms. UISES, an investment simulation, was designed by Oltheten and her colleague, Antoinette Tessmer. (See URL left.)

The RSE Program welcomes the involvement of interested educators and scientists at all levels.

RSE Program http://www.ncsa.uiuc.edu/Edu/RSE/RSEProjects.html **Geosciences/Weather Visualizer** http://www.atmos.uiuc.edu/geosciences/geosciences.html

CyberProf http://crome.ccsr.uiuc.edu/cyberprof/cyberprofmain.html **UISES** http://www.cba.uiuc.edu/uises.html

Umesh Thakkar is RSE program coordinator with the NCSA Education and Outreach Program.

Some Principal Learners of Internet Tools

by Sarah Thomas

Rhonda Anderson Photography

Despite snow 125 people turned out for an update on Champaign County Network (CCNet) at the Beckman Institute in mid-January. Among community leaders present were (counterclockwise) librarian Jan Ison; Ameritech executives Doug Schuemann and Bob Sharkey with NCSA Director Larry Smarr; CCNet Chair Lee O'Neill; and farmer John Reifsteck. Champaign County is tied for number one of 20 online communities in Illinois for most informative community Web exhibit. Nationwide it is among the top 10%. CCNet exhibits are accessible via the URL below right.

With the abundance of educational resources appearing on the World Wide Web and the escalation of schools connecting to the Internet, the NCSA Education and Outreach Group's Workshops and Training team, headed by Melissa Kelly, concentrated its efforts last fall on classes for school administrators. The administrators, including principals, district superintendents, technology coordinators, librarians, and teachers, play a vital role in planning for Internet connectivity.

The course is taught by NCSA, and credit is given by the Illinois Administrators Academy from the Regional Offices of Education in Rantoul and Charleston. Participants have attended from school districts in 18 different counties in Illinois.

The course has four parts: an introduction to electronic mail and Web browsers; researching Web-based resources and projects appropriate to the K–12 classroom and school administrators; studying the role of technology in school improvement; and exploring issues related to developing a technology plan for a school or district.

Kelly invited several administrators who have already implemented technology plans in their schools to speak to the classes. The guest speakers discussed their process of gaining competency with computers and how they are pursuing Internet connectivity, including the use of satellites. After hearing about two wireless connection projects at schools in Mahomet and Marshall, other schools in areas without local Internet access providers are considering implementing such options. The administrators are also closely following the emerging plan from the Illinois State Board of Education to establish technology hubs and help connect more schools in Illinois to the Internet.

> The guest speakers discussed their process of gaining competency with computers and how they are pursuing Internet connectivity.

In addition to the classes for administrators, workshops were given throughout the semester to preservice teachers from the UIUC College of Education and teachers and students from around central Illinois. The primary focus of these sessions was learning to use the World Wide Web efficiently, including learning search strategies to find educational resources.

Plans for spring 1996 workshops include continuing to reach school administrators from around the state and hosting a workshop for area teachers to develop Web-based curriculum units that they can use the following school year. In the summer of 1996, workshops will again be offered to teachers and students who have Internet connections and are ready to begin integrating the new tools and materials into their classrooms. Additionally a new project is underway with the Urbana Adult Education Program to provide a train-the-trainer model to support adult literacy training in Illinois through the use of technology. Representatives of 30 programs in the state will then serve as trainers for staff and volunteers in their respective programs. Both types of training will use the just-in-time concept—meeting the needs of each program as they come online.

As educators continue to mature in computer literacy, more advanced topics will be introduced. •

Sarah Thomas is a workshop and training assistant in the NCSA Education and Outreach Program.

Networking Technologies http://www.ncsa.uiuc.edu/edu/classroom/admin/

CCNet http://ww.ncsa.uiuc.edu/People/alaina/com_online

HDF: the Next Generation

by Mike Folk and Quincey Koziol

COMPUTATIONAL SCIENTISTS RARELY USE ONLY ONE COMPUTER. TYPICALLY THEY USE ONE OR MORE LARGE COMPUTERS TO RUN MODELS AND PROCESS DATA, POSSIBLY ANOTHER COMPUTER TO RENDER RESULTS FOR VISUALIZATION, AND STILL ANOTHER COMPUTER, ON A DESKTOP, TO FURTHER ANALYZE AND VISUALIZE THE DATA. THE NEED TO USE A MIX OF COMPUTERS AND TO TRANSPORT LARGE AMOUNTS OF DIFFERENT KINDS OF SCIENTIFIC DATA AMONG MANY DIFFERENT COMPUTERS WAS AN EARLY DATA MANAGEMENT PROBLEM FOR MANY OF NCSA'S SCIENTISTS.

In response to this use of several computers, NCSA developed the Hierarchical Data Format (HDF) in 1988. NCSA HDF is a portable, self-describing data format for moving and sharing scientific data in networked, heterogeneous computing environments. HDF can store several different kinds of data objects: multidimensional arrays, raster images, color palettes, and tables. It allows individual scientists to mix and group different kinds of data in one file, according to their needs.

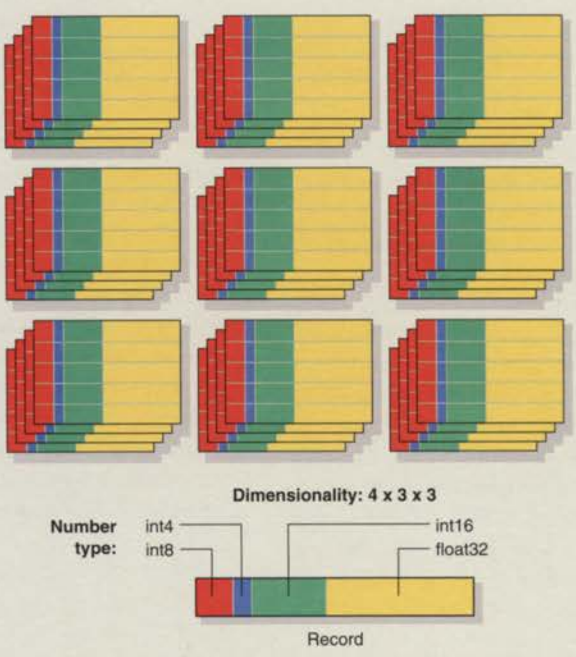
NCSA provides a library of application programming interfaces (APIs) for reading and writing HDF as well as workstation tools for visualizing data stored in HDF files. With the library and tools providing easy access to HDF, an enthusiastic user community emerged almost immediately. Users included NCSA's scientific community, but it also extended to other organizations and institutions, as well as several international users.

Early extensions to HDF: a second-generation HDF

HDF's designers understood that over time new, unforeseen requirements would emerge that HDF would not be able to handle and that it needed to be extendable. They designed HDF in such a way that new types of data could be added when original structures were insufficient.

Indeed soon after scientists began using HDF, they asked for enhancements: a table structure, a way of grouping objects within HDF files, support for data compression, and more data types. They also asked for changes in the HDF library, including more powerful APIs, user-defined attributes

Figure 1. Multidimensional array of elements.



Illustrations by Marshall Greenberg

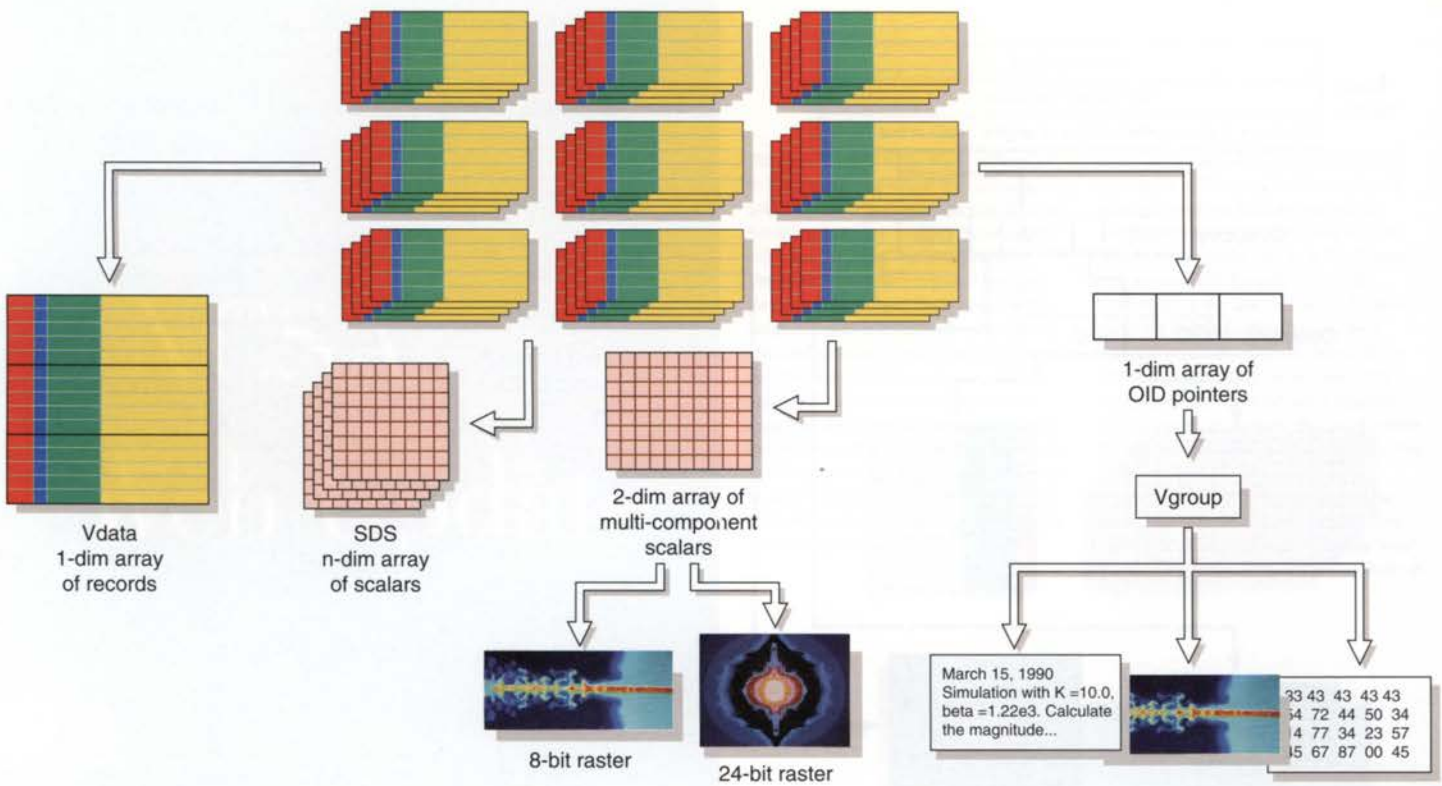


Figure 2. Derivation of existing HDF objects from new object type.

for HDF objects, and support within the library for another popular scientific data file format called netCDF.

From these needs a "second generation" HDF was created. The second generation of HDF is compatible with the first generation of HDF, but includes these new features:

- more powerful APIs (including the ability to access more than one file or data object at a time), user-defined attributes, and support for the netCDF data model
- an extended tag structure, making it possible to store data in a list of linked blocks or in an external file
- data compression for multidimensional arrays, support for many compression schemes and for new number types

The second, and current, generation of HDF is HDF 4.0.

New requirements: EOSDIS and Grand Challenges

Recently two new classes of HDF users have pushed the limits of the current implementation of HDF—the Grand Challenge and global change research communities.

Grand Challenge projects address problems in science and engineering whose solutions can be advanced by applying HPCC technologies. These problems involve very large datasets and typically run on fast, multiprocessing machines that require very fast I/O.

Global change research collects, organizes, and processes large amounts of data in order to understand how the Earth works. A fundamental component of global change research is the Earth Observing System (EOS), a space-based observing system with instruments that will ultimately collect terabytes of data daily. The EOS Data and Information System (EOSDIS) will use HDF as a standard format for storing much EOS data.

Some of these applications call for data files with thousands of data structures. Others store very large images, arrays, or tables. Some will have complex collections of interrelated data and metadata. These applications also frequently use computing technologies, such as object oriented approaches, to manage and manipulate data.

EOSDIS and Grand Challenge projects pose a whole new set of needs for HDF, needs that HDF 4.0 does not satisfy, including:

- a need to store very large objects (current HDF limit: 2 gigabytes)
- a need to store large numbers of objects (current limit: 20,000 objects)
- more general, flexible data models
- performance improvements
- compatibility with object-oriented databases and distributed object technologies

To address these new needs, the NCSA HDF project has support from NASA to design and build a prototype for the next generation of HDF—code name "BigHDF."

BigHDF: Accommodating variety

Current plans call for three fundamental changes in HDF: a unified data model, a simpler file structure, and a new I/O library.

New data model

HDF now supports several different objects, but the proposed data model will support only one: a multidimensional array of elements (see Figure 1 on page 14).

The new object will have two required attributes: dimensionality (the number and sizes of dimensions) and a data type (a definition of the array elements type). More data types will be supported, including complex numbers, date and time, pointers, and record structures.

Objects will include optional user-defined attributes of the form "parameter =

Mike Folk is the manager and research lead for NCSA HDF.
Quincey Koziol is a research programmer on the HDF project and design lead for NCSA BigHDF.

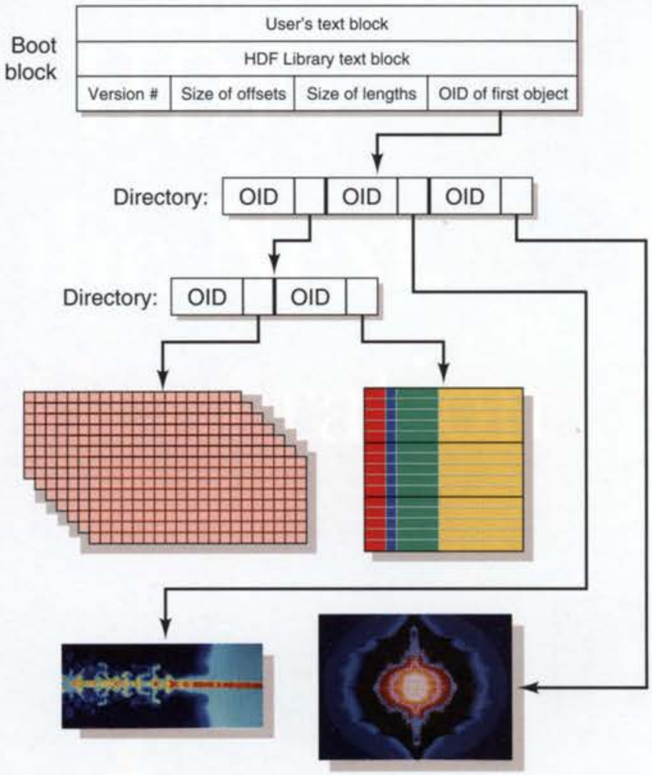


Figure 3. Hierarchical file structure.

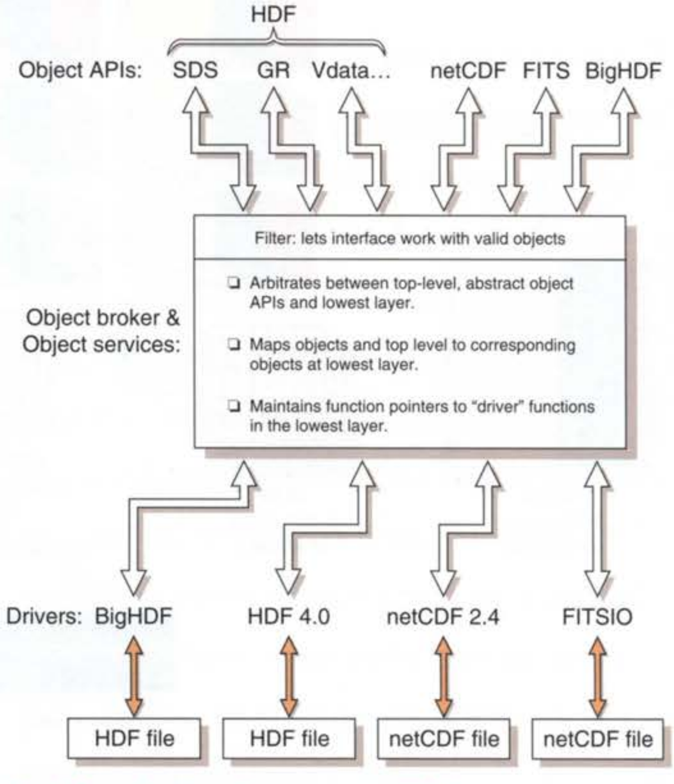


Figure 4. Proposed architecture for next generation HDF library.

value." Users will specify optional physical storage schemes for the data. By default, objects will be stored contiguously in a file, but alternative physical formats will be available, giving data producers some control over the physical organization of their data. Alternative physical formats will include external, linked-block, chunked, or compressed storage and an indexed structure.

For backward compatibility, the new HDF object is designed so that all current objects can be defined as subtypes of this basic object type (see Figure 2 on page 15).

New file structure

The new file structure will support files and objects of any size and any number of objects. The internal structure for describing objects is simpler than the current structure and should provide faster, easier access to objects. The three object parts are an object ID (OID), a header record with information required to describe the object, and the object itself.

The OID will be a large number, making it possible to represent a correspondingly large number of objects in an HDF file. The header record will contain or point to information about the object, including user-defined attributes and a description of the physical storage scheme used to store the object.

The entire file structure will consist of a boot block, followed by one or more objects. The boot block will begin with two text blocks that provide the user and the HDF library space for storing messages. It also will include the sizes of numbers that represent the offsets and lengths of objects, making it possible to support objects of virtually any size in the file. The last item in the boot block points to the first object in the file. If more than one object is in the file, the first object will be an array of OIDs. This kind of grouping can be repeated to create a hierarchical organization of objects within the file (see Figure 3 above).

Focus on interoperability

In planning the next generation HDF library, NCSA developers hope to exploit similarities between HDF and other popular scientific data formats by building a system that understands a variety of different data models and formats (see Figure 4 above). APIs at the top level allow programs to view data according to a variety of different data models. These APIs communicate with the middle layer through "object brokers" that rewrite their requests in terms of a common model. The middle layer also determines which service needs to be invoked to read or write the data and then invokes the necessary service.

The service layer consists of different file format drivers, each of which reads from or writes to one file format. Each driver has a well-documented interface for transferring objects and lists of objects to the higher arbitration layer. Possible drivers in the first implementation include HDF, BigHDF, netCDF, and FITS.

Prototyping planned for '96

NCSA's development team is eager for feedback on the BigHDF proposal (for detailed information see the URL at left). Turning the proposal's key features into a prototype is the goal for 1996.

BigHDF Proposal http://hdf.ncsa.uiuc.edu:8001/presentations/BigHDF/BigHDF.html

NCSA Web Update

The recent World Wide Web conference in Boston was the site of the inaugural presentation of the SoftQuad Web Award honoring someone "whose vision and work helped make the Web possible." The 1995 honor was bestowed upon Dr. Douglas C. Engelbart, the inventor of the graphical user interface, shared-screen teleconferencing, context-sensitive help, and the now-ubiquitous mouse. The biographical statement distributed about Engelbart notes that he "anticipated and helped shape the computing environment in which we live and work and continues to offer direction, particularly in the area of the World Wide Web—the most visible manifestation of his vision." Engelbart (seated middle) is shown signing copies of *Boosting Our Collective IQ*, selected readings from his more than thirty-year career in organizational computing. Also signing copies is Tim Berners-Lee (seated left), the author of HTML. Looking on is the late Yuri Rubinsky, sponsor of the SoftQuad Web Award and a member of the International World Wide Web Conference Committee. (Rubinsky, a writer, publisher, and software developer, died suddenly in January.) NCSA's Barbara Kucera (seated right), also a member of the conference committee, numbers the signed copies that will be sold to raise funds for the Bootstrap Institute founded by Engelbart. To order a signed copy, send email to info@bootstrap.org or call Mary at (510) 713-3550.

NCSA HTTPd Server Software Released

The Software Development Group released the latest version of the NCSA HTTPd server software in November 1995. Version 1.5 of the popular Web server software was a major rewrite of the code and includes multiple directory indexes, enhanced access control, internal image map support, Kerberos authentication, and virtual host support (for either multiple servers or a single server with multiple IP addresses). The documentation is also updated.

Setup is now even easier with the new "OneStep Downloader" page that walks a webmaster through the seven basic directives needed to configure a server for a site. Additional customization of the server daemon can take place after the initial set-up.

NCSA makes PC Magazine's Top 100 List

NCSA was named as a top 100 Web site by *PC Magazine* in December 1995. NCSA was named in the Internet Resources category. The online blurb about the center's Web server calls NCSA a "veteran site [that] has always been a hub for Internet resources." www.ncsa.uiuc.edu is cited for supplying Internet pointers, industry publications, multimedia exhibits, and "myriad" Internet software, including NCSA Mosaic.

Visit the exhibit at the URL listed right. The editors are looking for nominations for top sites—maybe your server would qualify!

NCSA Web Server Software Popular

Each year, Netcraft—a United Kingdom-based company—conducts a survey to determine what Web server software is being used on Internet-connected computers. Their November 1995 survey of more that 31,500 servers revealed that NCSA HTTPd software is used more frequently than similar software produced by other Internet software providers.

The NCSA HTTPd software is used by more than 14,000 servers—45% of all servers surveyed. NCSA's software commands a large market share for servers in the .gov domain (government) with almost 70% of the servers. NCSA HTTPd has 60% of the education domain (.edu) and 40% of the commercial (.com) market.

Vive la Web!

The Fifth International WWW Conference will be held May 6–11 at CNIT-Paris La Defense, located on the western side of Paris, France. A forum for discussion and dissemination of advances in Web technology, the conference was most recently held in Boston (December 1995). Heightened interest in the Web upped participant numbers from 382 at the first meeting at CERN in Geneva, Switzerland (May 1994) to nearly 3,000 nineteen months later in Boston. Each conference since the first has been "sold out" long before the registration deadline.

Innovations and new software for the Web are presented in technical papers and often become the processes used for organization and coordination of the conference itself. Examples of new Web technology that is now standard fare for the conference preparation include online paper review, registration, an online survey, and M-BONE live video over the Net.

The Web conferences will become annual events (rather than biannual) beginning in the spring of 1997, primarily due to the additional planning time needed because of the phenomenal increase in attendance. The site of the 1997 conference will be announced in February.

PC Magazine http://www.zdnet.com/~pcmag/special/web100/top100.htm **Netcraft** http://www.netcraft.com/Survey/
HTTPd Server Software ftp://ftp.ncsa.uiuc.edu/Web/httpd/Unix/ncsa_httpd/httpd_1.5 **WWW5** http://www.w3.org/pub/Conferences/WWW5/
OneStep Downloader http://hoohoo.ncsa.uiuc.edu/docs/setup/OneStep.html **Bootstrap Institute** http://www.bootstrap.org

NCSA's Second Industrial Partner User Meeting Addresses Security

by Allison Miller

NCSA Security Coordinator Ken Rowe has been known to say that the only way to ensure a 100% secure computer network is to keep it unplugged in a locked room. The impact of Rowe's recent remark on a business community that is increasingly submerged in electronic commerce was strong. A concern generated by those attempting to retain the right to privacy in an insecure environment was one of the driving forces behind NCSA's Second Industrial Partner User Meeting.

Attendees take a coffee break.

Mirroring the success of last year's Industrial Partner User Meeting [see *access*, Spring 1994, page 25], this year's meeting received positive feedback. Combined with a focus on security, the November 1995 Industrial Partner User Meeting at the Beckman Institute was attended by almost 40 participants.

Among the impressive panel put together by Rowe was Assistant U.S. Attorney General Colleen Coughlin, who addressed participants on computer crime and industrial espionage trends. Citing a 498% growth in computer intrusion since 1991, Coughlin went on to discuss some of the problems with existing laws governing industrial espionage and the possible threats to those conducting business via the Internet.

"The Internet expands everyone's vulnerability enormously. . . . If you are linked to the outer world, you are vulnerable to hackers," said Coughlin.

Aspects of civil liability and network law were discussed by William J. Cook of the Chicago law firm Willian, Brinks, Hofer, Gilson & Lione. Cook broached numerous legal cases reflecting the necessity to be alert to source data, network capabilities, and contract stipulations.

Moira West-Brown, manager of the Incident Handling Computer Emergency Response Team (CERT), discussed CERT's role in security and security infringement incident response. West-Brown also addressed the growing sophistication of computer crime exemplified by incidents over the last decade.

Deborah Cooper, chairperson of IEEE's Security and Privacy Technology Committee, discussed future directions for security technology. "You're never going to get a secure Internet or totally transparent security," said Cooper, who owns and founded a computer security business, Deborah M. Cooper Co.

In the afternoon participants separated into two tracks. Those in the User Meeting Track focused on questions and topics of interest to remote users of NCSA's facilities.

"The NCSA industrial user group meeting is for the people who are the day-to-day users of the center. The more they understand about NCSA's environment and people, the more comfortable they will feel," said Joe Blackmon, former NCSA industrial program manager.

Topics discussed in the User Meeting Track included a hardware/software update by Melanie Loots, associate director of NCSA's Applications Group; HKS ABAQUS by Paul Sorenson, vice president and partner of Hibbitt, Karlsson & Sorenson Inc.; AskMe Web-based user help system by John Towns, manager of NCSA's Consulting Services; and industrial queues by Mike Pflugmacher, member of NCSA's Advanced Computing Group.

Others attended the NCSA Security Track that provided an extension of the security material discussed in the morning sessions with a focus on NCSA's security efforts.

"The intent of the morning session was to raise concerns, to make sure the users and their security people understand the significance of the problems out there. It is a real issue that needs to be dealt with. The intent of the afternoon session was to make them feel less uncomfortable. When I say 'less uncomfortable,' I mean that I don't want anyone to feel comfortable. When you feel comfortable about security, you get in trouble," said Rowe.

The NCSA Security Track session included an overview of the NCSA security program by Rowe; a discussion of the role

Allison Miller, a senior in rhetoric at UIUC, is a student intern in the NCSA Marketing Communications Group.

Beckman Institute

of NCSA's Incident Response and Security Team by Jeff Rosendale, head of NCSA's Technology Management Group; an explanation of the NCSA security policy by Mike Carrillo, legal intern, NCSA Director's Office; a description of network security architecture by Randy Butler, technical program manager of NCSA's Networking and Security Group; and an explanation of the security features of the Andrew File System (AFS) by Jim Barlow, member of NCSA's Advanced Computing Group.

Combining the partner user meeting with a security meeting attracted a wide variety of participants, including users, scientists, executives, and security personnel. Attendees represented corporations including American Airlines, Caterpillar, Dow Chemical, Eastman Kodak Co., Eli Lilly & Co., FMC, J.P. Morgan, Motorola Inc., Phillips Petroleum Co., Schlumberger, Sears, Roebuck & Co., United Technologies, and the Tribune Co.

The conference highlighted the need to achieve an equilibrium between usability and security. "It's important to find the right balance between high-level security and the ease of use. Nothing is 100% secure. It's more a question of the level of security," said Blackmon. •

Photos by Jeff Carpenter, NCSA

Two New Partners for NCSA:

Sears

Sears, Roebuck and Co. recently joined NCSA's Industrial Program. Based in Hoffman Estates, IL, Sears is a leading retailer of apparel, home, and automotive products and services. The company operates 1,100 Sears department and specialty stores across the country.

"Sears is a national leader in seeking new information technologies to help them manage the ever-increasing competitive retail world," said NCSA Director Larry Smarr. "We feel NCSA can help them apply the evolving high-performance computing and information technologies to improve their ability to meet the growing expectations of distinct customer segments in increasingly fractured markets while holding the line on expense."

"We are excited about joining NCSA," said Joe Smialowski, Sears senior vice president and chief information officer. "NCSA has successfully guided its industrial partners in the use of leading-edge high-performance technologies. We plan to leverage NCSA's world class research talent and supercomputing facilities to solve key retailing challenges. We expect our partnership will yield important benefits to Sears in its drive to improve customer service and financial performance."

"Sears is a highly successful, high-profile retailing company, well suited to be our exclusive retail segment partner," said John Stevenson, NCSA's corporate officer and head of the Industrial Program. •

Allstate

NCSA recently announced that Allstate Insurance Co. joined the center's Industrial Program. Allstate Insurance Co., based in Northbrook, IL, is the nation's second-largest personal lines insurance company, insuring one of every eight homes and automobiles in the United States.

"We are excited about adding a world-leading insurance company to NCSA's Industrial Program," said Larry Smarr, director of NCSA.

"Allstate is looking forward to participating in the highly successful Industrial Program at NCSA," said Frank Pollard, senior vice president and chief information officer of Allstate. "We believe NCSA will give us a competitive edge by helping us apply emerging supercomputing, networking, software, and visualization tools to our analysis of customer and market databases."

"Allstate's commitment to maximize its capability to analyze its information resources is impressive," said John Stevenson, NCSA's corporate officer and head of the Industrial Program. "Their management is highly competitive and dedicated to improving their market position." •

NCSA's other industrial partners are AT&T; American Airlines; Caterpillar; Dow Chemical; Eastman Kodak Co.; Eli Lilly & Co.; FMC; J.P. Morgan; Motorola Inc.; Phillips Petroleum Co.; Schlumberger; Tribune Co.; and United Technologies.

Milestones

Courtesy National Academy of Engineering

NCSA Director Larry Smarr was inducted into the National Academy of Engineering at a ceremony in Washington, DC, in fall 1995. He was one of 77 elected to membership; among them was Steve Wallach of CONVEX Computer Corp. (right).

New staff

Carlton Bruett recently joined NCSA's Publications Group as media communications specialist. He is working with others in the Publications Group to coordinate NCSA's promotional image and is providing design support for producing a variety of NCSA publications in print and online. Bruett also serves as design liaison with other NCSA groups to ensure standardization of the center's image. Prior to joining NCSA, Bruett was the senior designer for Carle Foundation Hospital, Urbana, IL. He was educated at the University of Nebraska and UIUC. •

Jeff Carpenter , NCSA

NCSA's Software Development Group hosted Internet advocate Greg Elin on his coast-to-coast tour presented by *Total New York*. SDG Associate Director Joseph Hardin (top, center) greeted Elin on behalf of the center. Elin recorded his travels to Internet landmarks with a laptop computer, a digital video camera, and a cellular modem, which can be seen at http://www.totalny.com. America Online, Java, Sun Microsystems, Inc., and Sweet 'N Low sponsored the trip.

Awards

NCSA's **Adam Cain**, research programmer in SDG, was awarded first place for his tutorial at the fourth International World Wide Web Conference in December in Boston."Security, Authentication, and Privacy on the Web" was the title of his work.

Three NCSA publications won first place, Award of Excellence, at the 1995 Master Communicator Competition of the Central Illinois region. Judges from the Pittsburgh, PA, area chose the winners. IABC, PRSA, and WICI professional societies sponsored the annual event.

Winners were (1) 2nd International World Wide Web Conference program booklet, posters, and souvenir materials, including T-shirts and lapel pins (Design–Special Projects category), Manager of NCSA's Publications Group **Melissa Johnson**, Graphic Designer **Linda Jackson**, and Associate Editor **Paulette Sancken**, recipients; (2) "Introducing the World Wide Web" booklet (Writing–Technical/Interpretive Writing category), Publications Editor **Ginny Hudak-David**, recipient; and (3) *access* [July 1994, November 1994, and March 1995] (Publications–Magazine Publications category), **Johnson**, Publications Editor **Fran Bond**, **Jackson**, and Illustrator **John Havlik**, recipients. •

NCSA Technical Resources Catalog

The *NCSA Technical Resources Catalog* lists software manuals, preprints and technical reports, and other products available from NCSA and its vendors. It is available in hard copy as well as online on the Web and anonymous FTP. The 1996 catalog is enclosed with this issue of *access*, and printed addenda will be available upon request throughout the year. The online versions will be kept up to date to ensure timely ordering information. See the inside back cover of *access* for details. •

NCSA Technical Resources Catalog http://www.ncsa.uiuc.edu/Pubs/TechResCatalog/TRC.TOC.html

For Further Information

Documentation Orders

Articles in this *access* may refer to items that are available through the *NCSA Technical Resources Catalog*. To receive a hard copy of the catalog, send your request to Orders for Publications, NCSA Software, and Multimedia [see NCSA Contacts, page 2]. To view the catalog online, access the URL: http://www.ncsa.uiuc.edu/Pubs/TechResCatalog/TRC.TOC.html. To obtain the catalog from anonymous ftp, see instructions below. The catalog is in the /ncsapubs/TechResCatalog directory.

Accessing NCSA's Servers

Many NCSA publications (e.g., calendar of events, user guides, *access*, technical reports) as well as software are available via the Internet on one of two NCSA servers: anonymous FTP or the World Wide Web. If you are connected to the Internet, we encourage you to take advantage of the easy-to-use servers to copy or view files.

Anonymous FTP address:	ftp.ncsa.uiuc.edu
NCSA WWW Home Page:	http://www.ncsa.uiuc.edu/

NOTE: References in *access* to a URL refer to the server address and file location information used by Web browser software to retrieve documents.

If you have any questions about accessing the servers, contact your local system administrator or network expert. Instructions for accessing the anonymous FTP server follow.

Downloading from Anonymous FTP Server

A number of NCSA publications are installed on the NCSA anonymous FTP server. If you are connected to the Internet, you can download NCSA publications by following the procedures below. If you have any questions regarding the connection or procedure, consult your local system administrator or network expert.

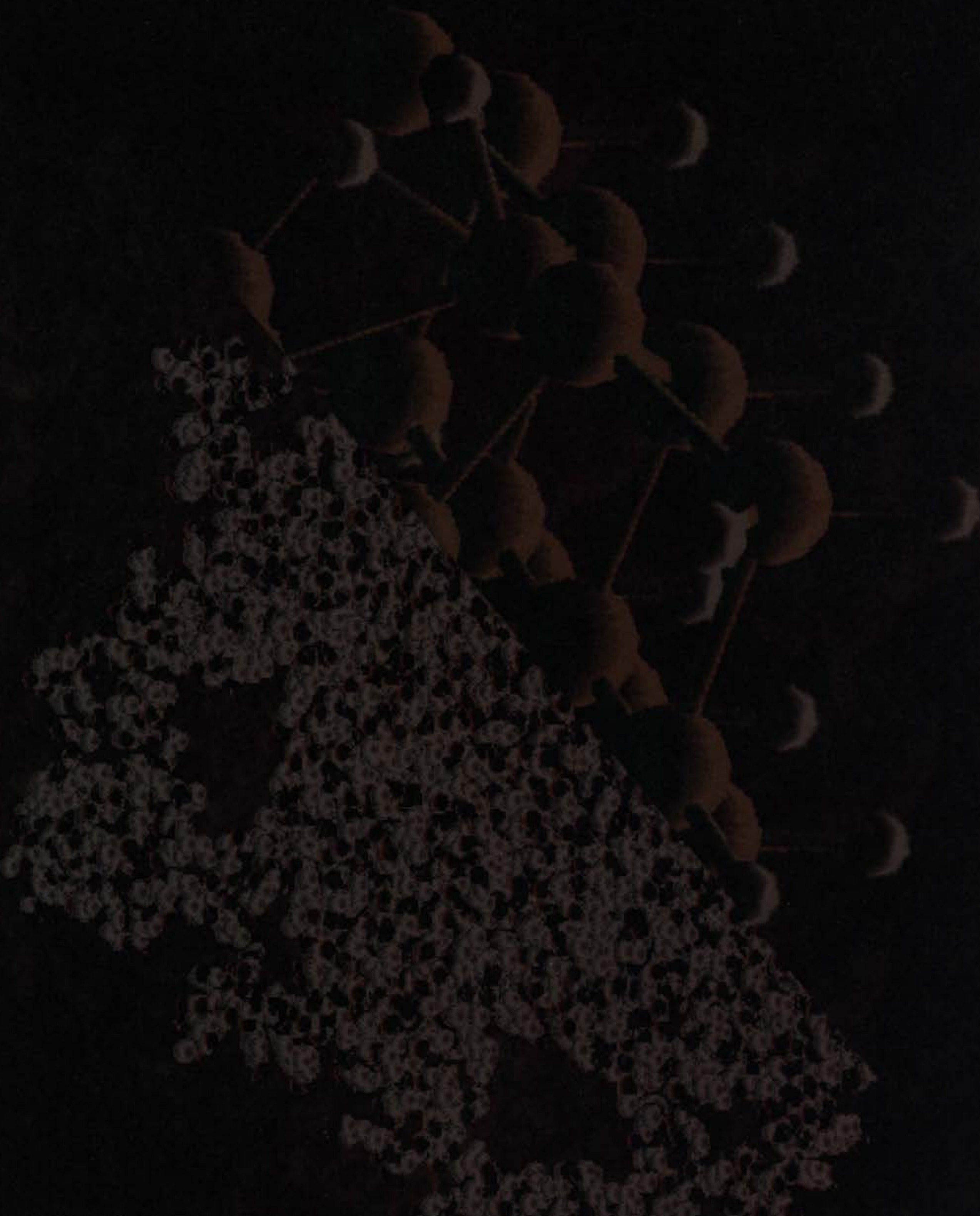
1. Log on to a host at your site that is connected to the Internet and running software supporting the FTP command.
2. Invoke FTP by entering the Internet address of the server: `ftp ftp.ncsa.uiuc.edu`
3. Log on using `anonymous` for the name.
4. Enter your local login name and address (e.g., smith@ncsa.uiuc.edu) for the password.
5. Enter `get README.FIRST` to transfer the instructions file (ASCII) to your local host.
6. Enter `quit` to exit FTP and return to your local host.
7. The NCSA publications are located in the `/ncsapubs` directory.

general abbreviations

ARPA	Advanced Research Projects Agency
CTC	Cornell Theory Center
EVL	Electronic Visualization Laboratory
HPCC	High Performance Computing and Communications
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCSA	National Center for Supercomputing Applications
NII	National Information Infrastructure
NSF	National Science Foundation
PSC	Pittsburgh Supercomputing Center
SDSC	San Diego Supercomputer Center
SGI	Silicon Graphics Inc.
TMC	Thinking Machines Corp.
UIC	University of Illinois at Chicago
UIUC	University of Illinois at Urbana-Champaign
URL	Uniform Resource Locator
Web	World Wide Web

NCSA abbreviations

Apps	Applications Group
CAVE	Cave Automatic Virtual Environment
C&C	Computing and Communications Group
E&O	Education and Outreach Program
F&A	Finance and Administration
IP	Industrial Program
MarComm	Marketing Communications Group
SDG	Software Development Group
VEG	Virtual Environments Group



NCSA

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
152 Computing Applications Building
605 East Springfield Avenue
Champaign, IL 61820-5518